OIPE

RAW SEQUENCE LISTING   DATE: 04/26/2001
PATENT APPLICATION: US/09/785,269   TIME: 10:29:41

Input Set : **N:\Crf3\RULE60\09785269.txt**
Output Set: **N:\CRF3\04262001\I785269.raw**

ENTERED

SEQUENCE LISTING
```
 4 (1) GENERAL INFORMATION:
 6      (i) APPLICANT: IMAI, Kensaku
 7                    KITAJIMA, Masato
 9     (ii) TITLE OF INVENTION: METHOD AND APPARATUS FOR AUTOMATICALLY
10                             REMOVING VECTOR UNIT IN DNA BASE SEQUENCE
12    (iii) NUMBER OF SEQUENCES: 19
14     (iv) CORRESPONDENCE ADDRESS:
15          (A) ADDRESSEE: Staas & Halsey
16          (B) STREET: 700 Eleventh Street, N.W., Suite 500
17          (C) CITY: Washington
18          (D) STATE: DC
19          (E) COUNTRY: US
20          (F) ZIP: 20001
22      (v) COMPUTER READABLE FORM:
23          (A) MEDIUM TYPE: Floppy disk
24          (B) COMPUTER: IBM PC compatible
25          (C) OPERATING SYSTEM: PC-DOS/MS-DOS
26          (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
28     (vi) CURRENT APPLICATION DATA:
C--> 29        (A) APPLICATION NUMBER: US/09/785,269
C--> 30        (B) FILING DATE: 20-Feb-2001
31          (C) CLASSIFICATION:
34    (vii) PRIOR APPLICATION DATA:
35          (A) APPLICATION NUMBER: US 08/684,674
36          (B) FILING DATE: 22-JUL-1996
39   (viii) ATTORNEY/AGENT INFORMATION:
40          (A) NAME: Herbert, William F.
41          (B) REGISTRATION NUMBER: 31,024
42          (C) REFERENCE/DOCKET NUMBER: 862.1335/WFH
44     (ix) TELECOMMUNICATION INFORMATION:
45          (A) TELEPHONE: 2024341500
46          (B) TELEFAX: 2024341501
49 (2) INFORMATION FOR SEQ ID NO: 1:
51      (i) SEQUENCE CHARACTERISTICS:
52          (A) LENGTH: 57
53          (B) TYPE: nucleic acid
54          (C) STRANDEDNESS: double
55          (D) TOPOLOGY: linear
57     (ii) MOLECULE TYPE: DNA (genomic)
62     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
64 AAGCTTGCAT GCCTGCAGGT CGACTCTAGA GGATCCCCGG GTACCGAGCT CGAATTC          57
66 (2) INFORMATION FOR SEQ ID NO: 2:
68      (i) SEQUENCE CHARACTERISTICS:
69          (A) LENGTH: 18
70          (B) TYPE: nucleic acid
71          (C) STRANDEDNESS: double
```

RAW SEQUENCE LISTING   DATE: 04/26/2001
PATENT APPLICATION: US/09/785,269   TIME: 10:29:41

Input Set : N:\Crf3\RULE60\09785269.txt
Output Set: N:\CRF3\04262001\I785269.raw

```
 72          (D) TOPOLOGY: linear
 74     (ii) MOLECULE TYPE: DNA (genomic)
 79     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
 81 TGCACTTGAA CGCATGCT                                            18
 83 (2) INFORMATION FOR SEQ ID NO: 3:
 85     (i) SEQUENCE CHARACTERISTICS:
 86          (A) LENGTH: 17
 87          (B) TYPE: nucleic acid
 88          (C) STRANDEDNESS: double
 89          (D) TOPOLOGY: linear
 91     (ii) MOLECULE TYPE: DNA (genomic)
 96     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
 98 TGCACTTGAA CGCTGCT                                             17
100 (2) INFORMATION FOR SEQ ID NO: 4:
102     (i) SEQUENCE CHARACTERISTICS:
103          (A) LENGTH: 17
104          (B) TYPE: nucleic acid
105          (C) STRANDEDNESS: double
106          (D) TOPOLOGY: linear
108     (ii) MOLECULE TYPE: DNA (genomic)
113     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
115 TGCACTTGAC GCATGCT                                             17
117 (2) INFORMATION FOR SEQ ID NO: 5:
119     (i) SEQUENCE CHARACTERISTICS:
120          (A) LENGTH: 17
121          (B) TYPE: nucleic acid
122          (C) STRANDEDNESS: double
123          (D) TOPOLOGY: linear
125     (ii) MOLECULE TYPE: DNA (genomic)
130     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
132 TGCACTTGAC GCATGCT                                             17
134 (2) INFORMATION FOR SEQ ID NO: 6:
136     (i) SEQUENCE CHARACTERISTICS:
137          (A) LENGTH: 17
138          (B) TYPE: nucleic acid
139          (C) STRANDEDNESS: double
140          (D) TOPOLOGY: linear
142     (ii) MOLECULE TYPE: DNA (genomic)
147     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:
149 TGCCTTGAAC GCATGCT                                             17
151 (2) INFORMATION FOR SEQ ID NO: 7:
153     (i) SEQUENCE CHARACTERISTICS:
154          (A) LENGTH: 2686
155          (B) TYPE: nucleic acid
156          (C) STRANDEDNESS: double
157          (D) TOPOLOGY: linear
159     (ii) MOLECULE TYPE: DNA (genomic)
164     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:
166 TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA    60
```

RAW SEQUENCE LISTING   DATE: 04/26/2001
PATENT APPLICATION: US/09/785,269   TIME: 10:29:41

Input Set : N:\Crf3\RULE60\09785269.txt
Output Set: N:\CRF3\04262001\I785269.raw

```
168 CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG   120
170 TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC   180
172 ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGGCGCC   240
174 ATTCGCCATT CAGGCTGCGC AACTGTTGGG AAGGGCGATC GGTGCGGGCC TCTTCGCTAT   300
176 TACGCCAGCT GGCGAAAGGG GGATGTGCTG CAAGGCGATT AAGTTGGGTA ACGCCAGGGT   360
178 TTTCCCAGTC ACGACGTTGT AAAACGACGG CCAGTGCCAA GCTTGCATGC CTGCAGGTCG   420
180 ACTCTAGAGG ATCCCCGGGT ACCGAGCTCG AATTCGTAAT CATGGTCATA GCTGTTTCCT   480
182 GTGTGAAATT GTTATCCGCT CACAATTCCA CACAACATAC GAGCCGGAAG CATAAAGTGT   540
184 AAAGCCTGGG GTGCCTAATG AGTGAGCTAA CTCACATTAA TTGCGTTGCG CTCACTGCCC   600
186 GCTTTCCAGT CGGGAAACCT GTCGTGCCAG CTGCATTAAT GAATCGGCCA ACGCGCGGGG   660
188 AGAGGCGGTT TGCGTATTGG GCGCTCTTCC GCTTCCTCGC TCACTGACTC GCTGCGCTCG   720
190 GTCGTTCGGC TGCGGCGAGC GGTATCAGCT CACTCAAAGG CGGTAATACG GTTATCCACA   780
192 GAATCAGGGG ATAACGCAGG AAAGAACATG TGAGCAAAAG GCCAGCAAAA GGCCAGGAAC   840
194 CGTAAAAAGG CCGCGTTGCT GGCGTTTTTC CATAGGCTCC GCCCCCCTGA CGAGCATCAC   900
196 AAAAATCGAC GCTCAAGTCA GAGGTGGCGA AACCCGACAG GACTATAAAG ATACCAGGCG   960
198 TTTCCCCCTG GAAGCTCCCT CGTGCGCTCT CCTGTTCCGA CCCTGCCGCT TACCGGATAC  1020
200 CTGTCCGCCT TTCTCCCTTC GGGAAGCGTG GCGCTTTCTC AAAGCTCACG CTGTAGGTAT  1080
202 CTCAGTTCGG TGTAGGTCGT TCGCTCCAAG CTGGGCTGTG TGCACGAACC CCCCGTTCAG  1140
204 CCCGACCGCT GCGCCTTATC CGGTAACTAT CGTCTTGAGT CCAACCCGGT AAGACACGAC  1200
206 TTATCGCCAC TGGCAGCAGC CACTGGTAAC AGGATTAGCA GAGCGAGGTA TGTAGGCGGT  1260
208 GCTACAGAGT TCTTGAAGTG GTGGCCTAAC TACGGCTACA CTAGAAGAAC AGTATTTGGT  1320
210 ATCTGCGCTC TGCTGAAGCC AGTTACCTTC GGAAAAAGAG TTGGTAGCTC TTGATCCGGC  1380
212 AAACAAACCA CCGCTGGTAG CGGTGGTTTT TTTGTTTGCA AGCAGCAGAT TACGCGCAGA  1440
214 AAAAAAGGAT CTCAAGAAGA TCCTTTGATC TTTTCTACGG GGTCTGACGC TCAGTGGAAC  1500
216 GAAAACTCAC GTTAAGGGAT TTTGGTCATG AGATTATCAA AAAGGATCTT CACCTAGATC  1560
218 CTTTTAAATT AAAAATGAAG TTTTAAATCA ATCTAAAGTA TATATGAGTA AACTTGGTCT  1620
220 GACAGTTACC AATGCTTAAT CAGTGAGGCA CCTATCTCAG CGATCTGTCT ATTTCGTTCA  1680
222 TCCATAGTTG CCTGACTCCC CGTCGTGTAG ATAACTACGA TACGGGAGGG CTTACCATCT  1740
224 GGCCCCAGTG CTGCAATGAT ACCGCGAGAC CCACGCTCAC CGGCTCCAGA TTTATCAGCA  1800
226 ATAAACCAGC CAGCCGGAAG GGCCGAGCGC AGAAGTGGTC CTGCAACTTT ATCCGCCTCC  1860
228 ATCCAGTCTA TTAATTGTTG CCGGGAAGCT AGAGTAAGTA GTTCGCCAGT TAATAGTTTG  1920
230 CGCAACGTTG TTGCCATTGC TACAGGCATC GTGGTGTCAC GCTCGTCGTT TGGTATGGCT  1980
232 TCATTCAGCT CCGGTTCCCA ACGATCAAGG CGAGTTACAT GATCCCCCAT GTTGTGCAAA  2040
234 AAAGCGGTTA GCTCCTTCGG TCCTCCGATC GTTGTCAGAA GTAAGTTGGC CGCAGTGTTA  2100
236 TCACTCATGG TTATGGCAGC ACTGCATAAT TCTCTTACTG TCATGCCATC CGTAAGATGC  2160
238 TTTTCTGTGA CTGGTGAGTA CTCAACCAAG TCATTCTGAG AATAGTGTAT GCGGCGACCG  2220
240 AGTTGCTCTT GCCCGGCGTC AATACGGGAT AATACCGCGC CACATAGCAG AACTTTAAAA  2280
242 GTGCTCATCA TTGGAAAACG TTCTTCGGGG CGAAAACTCT CAAGGATCTT ACCGCTGTTG  2340
244 AGATCCAGTT CGATGTAACC CACTCGTGCA CCCAACTGAT CTTCAGCATC TTTTACTTTC  2400
246 ACCAGCGTTT CTGGGTGAGC AAAAACAGGA AGGCAAAATG CCGCAAAAAA GGGAATAAGG  2460
248 GCGACACGGA AATGTTGAAT ACTCATACTC TTCCTTTTTC AATATTATTG AAGCATTTAT  2520
250 CAGGGTTATT GTCTCATGAG CGGATACATA TTTGAATGTA TTTAGAAAAA TAAACAAATA  2580
252 GGGGTTCCGC GCACATTTCC CCGAAAAGTG CCACCTGACG TCTAAGAAAC CATTATTATC  2640
254 ATGACATTAA CCTATAAAAA TAGGCGTATC ACGAGGCCCT TCGTC              2686
256 (2) INFORMATION FOR SEQ ID NO: 8:
258       (i) SEQUENCE CHARACTERISTICS:
259           (A) LENGTH: 66
260           (B) TYPE: nucleic acid
261           (C) STRANDEDNESS: double
```

RAW SEQUENCE LISTING   DATE: 04/26/2001
PATENT APPLICATION: US/09/785,269   TIME: 10:29:41

Input Set : N:\Crf3\RULE60\09785269.txt
Output Set: N:\CRF3\04262001\I785269.raw

```
262          (D) TOPOLOGY: linear
264     (ii) MOLECULE TYPE: DNA (genomic)
269     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:
271 GTGCCAAGCT TGCATGCCTG CAGGTCGACT CTAGAGGATC CCCGGTACCG AGCTCGAATT        60
273 CGTAAT                                                                  66
275 (2) INFORMATION FOR SEQ ID NO: 9:
277     (i) SEQUENCE CHARACTERISTICS:
278          (A) LENGTH: 6
279          (B) TYPE: nucleic acid
280          (C) STRANDEDNESS: double
281          (D) TOPOLOGY: linear
283     (ii) MOLECULE TYPE: DNA (genomic)
288     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:
290 AAGCTT                                                                   6
292 (2) INFORMATION FOR SEQ ID NO: 10:
294     (i) SEQUENCE CHARACTERISTICS:
295         ·(A) LENGTH: 6
296          (B) TYPE: nucleic acid
297          (C) STRANDEDNESS: double
298          (D) TOPOLOGY: linear
300     (ii) MOLECULE TYPE: DNA (genomic)
305     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:
307 GCATGC                                                                   6
309 (2) INFORMATION FOR SEQ ID NO: 11:
311     (i) SEQUENCE CHARACTERISTICS:
312          (A) LENGTH: 6
313          (B) TYPE: nucleic acid
314          (C) STRANDEDNESS: double
315          (D) TOPOLOGY: linear
317     (ii) MOLECULE TYPE: DNA (genomic)
322     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:
324 CTGCAG                                                                   6
326 (2) INFORMATION FOR SEQ ID NO: 12:
328     (i) SEQUENCE CHARACTERISTICS:
329          (A) LENGTH: 6
330          (B) TYPE: nucleic acid
331          (C) STRANDEDNESS: double
332          (D) TOPOLOGY: linear
334     (ii) MOLECULE TYPE: DNA (genomic)
339     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:
341 GGTACC                                                                   6
343 (2) INFORMATION FOR SEQ ID NO: 13:
345     (i) SEQUENCE CHARACTERISTICS:
346          (A) LENGTH: 6
347          (B) TYPE: nucleic acid
348          (C) STRANDEDNESS: double
349          (D) TOPOLOGY: linear
351     (ii) MOLECULE TYPE: DNA (genomic)
356     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:
```

```
358 TCTAGA                                                          6
360 (2) INFORMATION FOR SEQ ID NO: 14:
362      (i) SEQUENCE CHARACTERISTICS:
363           (A) LENGTH: 6
364           (B) TYPE: nucleic acid
365           (C) STRANDEDNESS: double
366           (D) TOPOLOGY: linear
368      (ii) MOLECULE TYPE: DNA (genomic)
373      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:
375 GTCGAC                                                          6
377 (2) INFORMATION FOR SEQ ID NO: 15:
379      (i) SEQUENCE CHARACTERISTICS:
380           (A) LENGTH: 6
381           (B) TYPE: nucleic acid
382           (C) STRANDEDNESS: double
383           (D) TOPOLOGY: linear
385      (ii) MOLECULE TYPE: DNA (genomic)
390      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:
392 GTCGAC                                                          6
394 (2) INFORMATION FOR SEQ ID NO: 16:
396      (i) SEQUENCE CHARACTERISTICS:
397           (A) LENGTH: 6
398           (B) TYPE: nucleic acid
399           (C) STRANDEDNESS: double
400           (D) TOPOLOGY: linear
402      (ii) MOLECULE TYPE: DNA (genomic)
407      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:
409 CCCGGG                                                          6
411 (2) INFORMATION FOR SEQ ID NO: 17:
413      (i) SEQUENCE CHARACTERISTICS:
414           (A) LENGTH: 6
415           (B) TYPE: nucleic acid
416           (C) STRANDEDNESS: double
417           (D) TOPOLOGY: linear
419      (ii) MOLECULE TYPE: DNA (genomic)
424      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:
426 GAATTC                                                          6
428 (2) INFORMATION FOR SEQ ID NO: 18:
430      (i) SEQUENCE CHARACTERISTICS:
431           (A) LENGTH: 6
432           (B) TYPE: nucleic acid
433           (C) STRANDEDNESS: double
434           (D) TOPOLOGY: linear
436      (ii) MOLECULE TYPE: DNA (genomic)
441      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:
443 CCCGGG                                                          6
445 (2) INFORMATION FOR SEQ ID NO: 19:
447      (i) SEQUENCE CHARACTERISTICS:
448           (A) LENGTH: 6
```

VERIFICATION SUMMARY   DATE: 04/26/2001
PATENT APPLICATION: US/09/785,269   TIME: 10:29:42

Input Set : N:\Crf3\RULE60\09785269.txt
Output Set: N:\CRF3\04262001\I785269.raw

L:29 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:30 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]